

## BAB 2

### TINJAUAN PUSTAKA DAN DASAR TEORI

#### 2.1 Tinjauan Pustaka

Penelitian ini menggunakan beberapa sumber pustaka yang berhubungan dengan kasus yang akan diteliti. Pemanfaatan algoritma *apriori* sudah cukup banyak digunakan, antara lain dalam penelitiannya Lukman (2015) menggunakan algoritma *apriori* untuk menentukan penempatan rak barang. Andriyanto (2015) dengan penelitiannya untuk mengetahui pola penyediaan stok obat di Puskesmas Banguntapan II.

Menurut Anggraeni dkk (2014) analisis ini juga berguna untuk menyajikan informasi hubungan pembelian obat dengan melihat dari nilai *support* dan *confidence* tertinggi data transaksi yang diuji digunakan dari bulan Januari sampai Maret 2014.

Penelitian lainnya tentang algoritma *apriori* adalah tentang pemanfaatan *data* transaksi penjualan yang berguna untuk mengatur penempatan barang yang dilakukan oleh Agung dkk (2015). Dimana aplikasi dibangun untuk menyajikan informasi yang berguna untuk meningkatkan pendapatan toko dengan cara mengatur penempatan barang yang saling berhubungan satu sama lain ketika konsumen membeli.

Wahyudi dkk (2015) juga telah melakukan penelitian yang memanfaatkan algoritma *apriori* untuk menganalisis pola pembelian konsumen di Pamella supermarket. Aplikasi ini dibangun untuk mengetahui pola pembelian yang menghasilkan dari ketiga cabang supermarket pola pembeliannya tidak jauh berbeda.

Aplikasi yang dibuat oleh penulis ini berguna untuk menerapkan algoritma *apriori* untuk mengetahui pola pembelian konsumen untuk membantu pemilik swalayan dalam pengaturan *display* barang.

Tabel 2.1 Tinjauan Pustaka

No	Penulis	Objek	Metode	Hasil
1.	Rifa Lukman (2015)	Radjamart	Algoritma Apriori	Menemukan pola hubungan antar barang dan mengetahui letak barang yang direkomendasikan
2.	Andriyanto (2015)	Puskesmas Banguntapan II	Algoritma Apriori	Menghasilkan hubungan antar barang guna untuk mengetahui penyediaan stok obat dan semakin tinggi dan bervariasi juga nilai <i>support</i> , nilai <i>confidencenya</i> dan gabungan yang dapat dihasilkan.
3.	Hapsari Dita Anggraeni, Ragil Saputra, dan Beta Noranita (2014)	Apotek Setya Sehat Semarang	Algoritma Apriori	Nilai minimum support = 6 dan minimum confidence = 40% merupakan nilai yang menghasilkan aturan paling kuat pada bulan februari (1 aturan) dan Maret (2 aturan), sedangkan nilai minimum support = 6 dan minimum confidence 50% menghasilkan aturan paling kuat pada bulan Januari (3 aturan)

Tabel 2.1 Tinjauan Pustaka

No	Nama	Objek	Metode	Hasil
4.	Muhammad Thoriq Agung dan Bowo Nurhadiyono (2015)	Toko Tombo Ati	Algoritma Apriori	Pengaturan tata letak barang dilihat dari nilai confidence yang paling tinggi kemudian akan diletakan berdekatan, sedangkann dengan nilai support tertinggi akan diletakan ditempat yang paling mudah terlihat karena barang tersebut paling sering dibeli.
5.	M. Didik R. Wahyudi dan Fusna Failasufa (2015)	Pamella Supermarket	Algoritma Apriori	Pola pembelian dari ketiga cabang tidak jauh beda.
6.	Tri Wahyuningsih (2017)	Aswalayan	Algoritma Apriori	Menghasilkan hubungan barang yang sering dibeli sehingga bisa membantu pemilik dalam mengatur <i>display</i> barang.

## 2.2 Dasar Teori

### 2.2.1 Data mining

*Data mining* adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. Definisi lain diantaranya adalah pembelajaran berbasis induksi (*induction-based learning*) adalah proses pembentukan definisi – definisi konsep umum yang dilakukan dengan cara mengobservasi contoh – contoh spesifik dari konsep – konsep yang akan dipelajari (Hermawati, 2009).

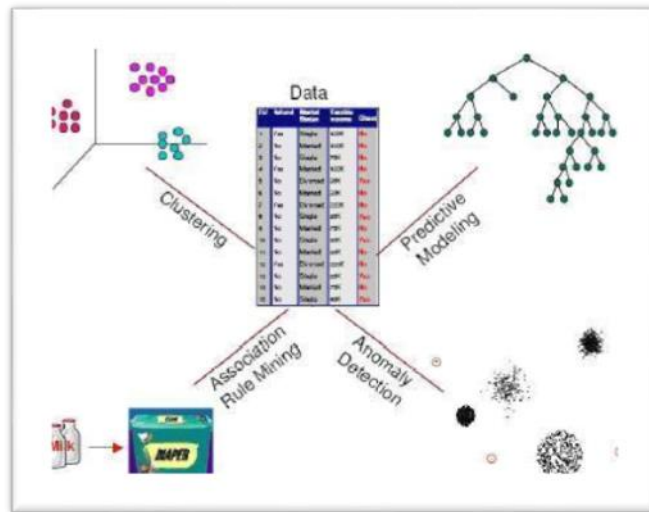
*Data mining* , sering juga disebut *knowledge discovery in databae* (KDD), adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Keluaran dari *data mining* ini bisa dipakai untuk memperbaiki pengambilan keputusan di masa depan (Santosa, 2007).

*Data mining* berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola – pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lainnya (Hermawati,2009).

Berdasarkan beberapa pengertian tersebut dapat ditarik kesimpulan bahwa *data mining* adalah suatu teknik menggali informasi berharga yang terpendam atau tersembunyi pada suatu koleksi data (database) yang sangat besar sehingga ditemukan suatu pola yang menarik yang sebelumnya tidak diketahui.

### **2.2.2 Metode Dalam Data Mining**

Menurut Prasetyo 2014, *data mining* dibagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan. Ilustrasi keempat pekerjaan tersebut dapat dilihat pada Gambar 2.1.



Gambar 2.1 Metode Dalam Data Mining

a. Model prediksi (*prediction modelling*)

Pekerjaan ini berkaitan dengan pembuatan sebuah model yang dapat melakukan pemetaan dari setiap himpunan variabel ke setiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan baru yang didapat. Ada 2 jenis model prediksi, yaitu klasifikasi dan regresi. Klasifikasi digunakan untuk variabel target diskret, sedangkan regresi digunakan untuk variabel target kontinu (Prasetyo, 2014).

b. Analisis cluster (*cluster analysis*)

Analisis kelompok melakukan pengelompokan data ke dalam sejumlah kelompok berdasarkan kesamaan karakteristik masing-masing data pada kelompok-kelompok yang ada. Data-data yang masuk dalam batas kesamaan dengan kelompoknya akan bergabung dalam kelompok tersebut, dan akan terpisah dalam kelompok yang berbeda jika keluar dari batas kesamaan kelompok tersebut (Prasetyo, 2014).

c. Analisis asosiasi (*association analysis*)

Analisis asosiasi digunakan untuk menemukan pola yang menggambarkan kekuatan hubungan fitur dalam data. Pola yang ditemukan biasanya merepresentasikan bentuk aturan implikasi atau subset fitur. Tujuannya adalah untuk menemukan pola yang menarik dengan cara yang efisien (Prasetyo, 2014).

d. Deteksi anomali (*anomaly detection*)

Pekerjaan deteksi anomali berkaitan dengan pengamatan sebuah data dari sejumlah data yang secara signifikan mempunyai karakteristik yang berbeda dari sisa data yang lain. Data-data yang karakteristiknya menyimpang (berbeda) dari data yang lain disebut sebagai *outlier*. Algoritma deteksi anomali yang baik harus mempunyai laju deteksi yang tinggi dan laju kesalahan yang rendah (Prasetyo, 2014).

### 2.2.3 Tahap – Tahap dalam Data Mining

Dalam *data mining* terdapat beberapa tahapan, sebagai berikut (Kusrini dan Luthfi, 2009) :

a. Pembersihan data (*data cleaning*)

Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan. Pada umumnya data yang diperoleh dari basis data suatu perusahaan, memiliki isian - isian yang tidak sempurna seperti data yang hilang, data yang tidak valid atau juga sekedar salah ketik. Selain itu, ada juga atribut – atribut data yang tidak relevan dengan hipotesa *data mining* yang dimiliki.

b. Integrasi data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai basis data ke dalam suatu basis data baru. Data yang diperlukan untuk *data mining* tidak hanya berasal dari satu basis data tetapi juga berasal dari beberapa basis data. Integrasi data dilakukan pada atribut – atribut yang mengidentifikasikan entitas – entitas yang unik seperti atribut nama, jenis produk, nomor pelanggan, dan lainnya.

c. Seleksi data (*data selection*)

Data yang ada pada basis data seringkali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk proses analisis yang akan diambil dari basis data. Sebagai contoh, sebuah kasus yang meneliti faktor kecenderungan orang membeli dalam kasus analisis keranjang belanja, tidak perlu mengambil nama pelanggan, cukup dengan id pelanggan.

d. Transformasi data (*data transformation*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*. Beberapa metode *data mining* membutuhkan format data yang khusus sebelum bisa diaplikasikan.

e. Proses mining

Proses *mining* merupakan proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

f. Evaluasi pola (*pattern evaluation*)

Evaluasi pola bertujuan untuk menemukan pola – pola menarik ke dalam basis pengetahuan yang ditemukan. Dalam tahap ini hasil dari teknik *data mining*

berupa pola – pola yang khas maupun moden prediksi dievaluasi untuk menilai apakah hipotesis yang ada memang tercapai.

g. Presentasi pengetahuan (*knowledge presentation*)

Presentasi pengetahuan merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

#### 2.2.4 Algoritma Apriori

Algoritma *apriori* termasuk jenis aturan asosiasi pada *data mining*. Selain *apriori*, yang termasuk pada golongan ini adalah metode *Generalized Rule Induction* dan *Algoritma Hash Based*. Aturan yang menyatakan asosiasi antara beberapa atribut sering disebut *affinity analysis* atau *market basket analysis* (Kusrini dkk, 2009).

Algoritma *apriori* merupakan salah satu metode untuk menggali kaidah asosiasi yang paling sederhana dan paling terkenal untuk menemukan pola frekuensi tinggi. Pola frekuensi tinggi adalah pola-pola item di dalam suatu database yang memiliki frekuensi atau *support* di atas ambang batas tertentu yang disebut dengan istilah *minimum support*. Pola frekuensi tinggi ini digunakan untuk menyusun aturan asosiatif dan juga beberapa teknik *data mining* lainnya.

Analisis asosiasi atau *association rule mining* adalah teknik *data mining* untuk menemukan aturan asosiatif antara suatu kombinasi item.

Menurut Han, dkk (2001), terdapat dua ukuran kepercayaan yang menunjukkan kepastian dan tingkat kegunaan suatu *rule* yang ditemukan yaitu :



### 1. *Support*

*Support* (dukungan) merupakan suatu ukuran yang menunjukkan seberapa besar dominasi suatu item atau *itemset* dari keseluruhan transaksi.

### 2. *Confidence*

*Confidence* (tingkat kepercayaan) adalah suatu ukuran yang menunjukkan hubungan antar item secara *conditional* (misalnya seberapa sering item B dibeli jika orang membeli item A).

Pada umumnya *association rule* yang didapatkan menarik apabila rule tersebut memenuhi baik minimum *support* maupun minimum *confidence* yang telah ditentukan oleh user.

Dalam Kusrini dkk. 2009, metodologi dasar analisis asosiasi terbagi menjadi 2 tahap :

#### 1. Analisis pola frekuensi tinggi

Tahap ini mencari kombinasi *item* yang memenuhi syarat *minimum* dari nilai *support* dalam *database*. Nilai *support* sebuah *item* diperoleh dengan mengikuti persamaan (2.1) (Kusrini, dkk. 2009).

$$Support(A) = \frac{\text{Jumlah transaksi mengandung } A}{\text{Total transaksi}} \times 100\% \quad (2.1)$$

Sedangkan nilai *support* dari 2 *item* diperoleh dari persamaan (2.2) dan persamaan (2.3) (Kusrini, dkk. 2009).

$$Support(A, B) = P(A \cap B) \quad (2.2)$$

$$Support(A \cap B) = \frac{\sum \text{Transaksi mengandung } A \text{ dan } B}{\sum \text{Transaksi}} \times 100\% \quad (2.3)$$

## 2. Pembentukan aturan asosiatif

Setelah semua pola frekuensi tinggi ditemukan, barulah dicari aturan asosiatif yang memenuhi syarat *minimum* untuk *confidence* dengan menghitung *confidence* aturan asosiatif  $A \rightarrow B$ . Nilai *confidence* dari aturan  $A \rightarrow B$  diperoleh dengan persamaan (2.4) (Kusrini, dkk. 2009).

$$Confidence = P(B | A) = \frac{\text{Jumlah transaksi mengandung A dan B}}{\text{Jumlah transaksi mengandung A}} \times 100\% \quad (2.4)$$

Algoritma *apriori* menurut Kusrini, dkk (2009) sebagai berikut :

1. Tentukan nilai minimum *frequent item set* yang menunjukkan *itemset* memiliki frekuensi kemunculan lebih dari nilai minimum yang telah ditentukan ( $\Phi$ ) untuk mendapatkan kandidat 1 *itemset*.
2. Setelah mendapatkan hasil dari iterasi ke 1, maka akan dilanjutkan untuk mencari kandidat 2 *itemset* dan selanjutnya. Proses akan terhenti ketika tidak menemukan yang sesuai dengan nilai minimum yang telah ditentukan.
3. Dari kandidat *itemset* yang telah diperoleh kemudian dihitung nilai *support* dan *confidence*-nya, dengan syarat harus memenuhi nilai minimum *support* dan *confidence* yang telah ditentukan.
4. Diperoleh *rules* yang dapat digunakan sebagai informasi oleh pengguna.